干旱区地理
ARID LAND GEOGRAPHY

自动土壤水分观测资料异常数据检测方法

李翠娜[1]， 刘天琦[2]， 吴东丽[1]
（1. 中国气象局气象探测中心，北京 100081；2. 内蒙古自治区气象信息中心，内蒙古 呼和浩特 010051）

摘 要：针对全国自动土壤水分实时小时观测资料，结合仪器标定方法、土壤水文物理常数等因素，研究土壤水分固有变化特征，分析异常数据误差来源和阈值判定，利用频率检测、界限值检测、无降水突变检测、异常峰值检测、恒值检测五类方法对土壤水分观测资料进行质量控制实验和检验。结果表明：(1) 自动土壤水分异常数据主要分为：粗大值、突变、异常峰值和僵值，主要由仪器失灵、土壤水文物理常数测定不准、传感器标校不合理等原因引起。(2) 频率检测可有效检出因仪器故障引起的错误数据。目前该方法已应用于综合气象观测数据质量控制业务系统，用于开展全国实时自动土壤水分小时数据的质量控制和质量评估。

关 键 词：自动土壤水分；质量控制；异常数据；仪器原理
文章编号：

土壤水分是研究土壤侵蚀、农作物旱情监测和产量预测等的重要指标，也是气候模型、水文模型、生态模型以及陆面过程模型的关键参数之一，监测土壤水分对农业、旱情、气候等具有重要意义[1-6]。频域反射(Frequency domain reflectomentry，FDR)技术以其高精度、高时空分辨率、实时、全天候等特点成为目前最具发展潜力的地基土壤水分监测技术之一。我国气象部门的自动土壤水分观测站网就是基于该方法建立的。土壤水分在观测过程中，存在多种因素影响着FDR型传感器测量的准确性，其误差来源包括温度、土壤质地、标定方法、设备自身稳定性和维护等，影响了土壤水分资料的业务应用。为充分发挥站网数据在干旱监测预警、验证卫星产品精度和模式模拟效果中的作用，需开展针对数据质量影响因子及处理算法的研究。

不同于常规气象观测数据，土壤水分观测数据地域空间分布极不均匀，受土壤质地和结构、标定方程等的影响比较大，目前还没有形成规范的业务化实时数据质量控制系统。近几年，许多学者致力于研究土壤水分数据质量自动控制技术。张志富等[7]通过对长序列人工观测土壤水分数据极端值的统计分析，研制了适用于全国自动土壤水分自动站逐小时数据的质量控制方案。王晓东等[8]通过对人工和自动土壤水分数据间绝对误差及相关性分析，探讨了导致自动土壤水分数据误差的主要原因，初步确定了安徽省自动土壤水分数据业务质量控制的方法。王良宇等[9]利用土壤水分自动数据间的相互关系，通过分析监测土壤各种参数特点，结合土壤水分再分配规律，提出了一种土壤水分数据质量控制预警方法。吴东丽等[10]利用范围检查、时变检查和持续性检查方法，对0~10 cm土壤水分数据进行了检查和分析。王佳强等[11]引入地温、降水的关系，从界限值检查、内部一致性检查、时间一致性检查等方面提出了一套适用于全国自动土壤水分观测资料的质量控制方法。上述方法多采用设定阈值的方法，虽然较为简单，但能有效识别出阈值内的异常值。而Dorigo等[12]采用谱分析方法，对欧洲国际土壤湿度网自动气象站数据进行质量控制分析，较为系统地建立了筛选剔除异常值的新方法。此后，各国学者利用该方法开展了大量研究。以上

收稿日期：2020-01-03； **修订日期**：2020-11-09
基金项目：国家重点研发专项分级质控和多源资料快速融合分析技术(2018YFC1507502)；自动土壤水分观测数据质量控制算法研究(AMF201603)资助
作者简介：李翠娜(1981-)，女，高级工程师，博士，主要从事多源观测质控校验关键技术研究与应用. E-mail: licn@cma.gov.cn

研究成果为本文进行土壤水分资料预处理和质量控制方案设计奠定了基础。

目前,国内外广泛开展的土壤水分数据质量控制的研究对象基本都围绕面向服务需求的相对湿度进行。但是,在土壤水分观测过程中,土壤水文物理特性测定值的准确性、标定方程的适用性、仪器安装的合理性等问题都会影响数据的准确性。利用相对湿度作为土壤水分质量控制对象,不能客观反映出仪器状态好坏或数据质量正确与否。为了提高土壤水分观测资料的应用效益,应从影响土壤水分数据质量的主要因素着手建立以频率和土壤体积含水量为主,相对湿度等计算值为辅的数据质量控制方法尤为重要。

本文以解决土壤水分观测数据存在的突出问题为质量控制的首要目标,借鉴国外ISMNQC方案[12],基于FDR型土壤水分传感器特殊的工作原理和方法,以设备原始输出频率值和体积含水量为对象,结合邻近自动气象站小时降水和地温数据,研究土壤水分数据固有变化规律,分析各种原因引起的错误数据表现形式和判定阈值,初步设计和研制了一套较为实用的适用于小时土壤水分观测的异常数据检测方法,以期进一步提高土壤水分数据质量,为农业气象干旱监测、预报工作提供可信度较高的监测信息资料。

1 资料与方法

自动土壤水分传感器采用频域反射法(FDR)监测,基于LC振荡电路,根据电磁波在介质中传播速度来测试介质的介电常数,从而测得土壤水分。观测数据分为器测值和导出值,器测值是传感器直接测得的频率或电压;体积含水量是基于器测值结合传感器实验室标定和田间标定的参数方程计算而得的一级数据;土壤重量含水率、相对湿度和土壤有效水分贮存量是基于一级数据,通过人工测定的土壤水文物理常数换算而得的二级数据。

自动土壤水分观测深度共有8层,分别是0~10 cm、10~20 cm、20~30 cm、30~40 cm、40~50 cm、50~60 cm、70~80 cm和90~100 cm。实际观测中可根据需求,结合土壤质地、地下水位、母质层深度等进行观测深度调整[13-14]。

本研究选用2013—2015年全国2000余部自动土壤水分观测站资料,设计土壤水分观测资料异常数据检测方法,同时选用2019年6—9月全国自动土壤水分观测站资料对检测方法的应用效果进行检验。同时段的逐时降水量、土壤温度等资料来自距离被检土壤水分观测站5 km范围内最近的地面国家级自动站或区域站。

2 异常数据检测方法与流程

2.1 土壤体积含水量变化特征

降水是影响土壤体积含水量的重要因素,发生降水后,土壤体积含水量升高(以30 cm为例),随着降水对土壤水分的入渗补给,到达峰值之后迅速下降,土壤由湿变干,最终达到稳定的特征曲线(图1a~b);由于冰的介电常数明显低于水的介电常数,土壤水分冻结时,土壤体积含水量明显低于正常值。同时土壤体积含水量受温度影响明显[15],土壤冻融过程中,土壤体积含水量随温度降低而减小,随温度升高而增大(图1c);土壤表层(0~10 cm)是整个土壤圈与外界自然环境的接触面,直接受到降水、蒸发等外界环境影响,相较于深层土壤,土壤体积含水量变异系数大[16]。除土壤表层外,其余各层土壤体积含水量之间具有高相关性,且相邻层次土壤体积含水量呈极显著相关,相互作用明显[17-19](图1d)。

2.2 异常数据误差来源分析

由于传感器故障,在无降水情况或24 h降水量小于某一临界值时,前后相邻时次土壤水分发生突变(图2a);土壤龟裂时,土壤水分在一段时间内观测值持续下降直至接近甚至达到0(图2b),龟裂多发生在黏性土壤[20];连续发生降水时土壤体积含水量长时间保持不变或发生微小变化(图2c);传感器标校不合理或观测地点的观测值缺乏代表性导致土壤水分测量值和真实值会存在系统误差[21],通过适当的传感器标校土壤水分值恢复正常(图2d)。

2.3 检测方法

2.3.1 基于频率特征的仪器故障检测方法(FD)

根据土壤水分传感器观测原理,土壤中测的频率应介于空气频率和水中频率之间。通过研究分析其变化特征确定界限值,从而检测仪器故障引起的数据错误。

$$f_{tD\min} \leqslant f_{tD} \leqslant f_{tD\max} \tag{1}$$

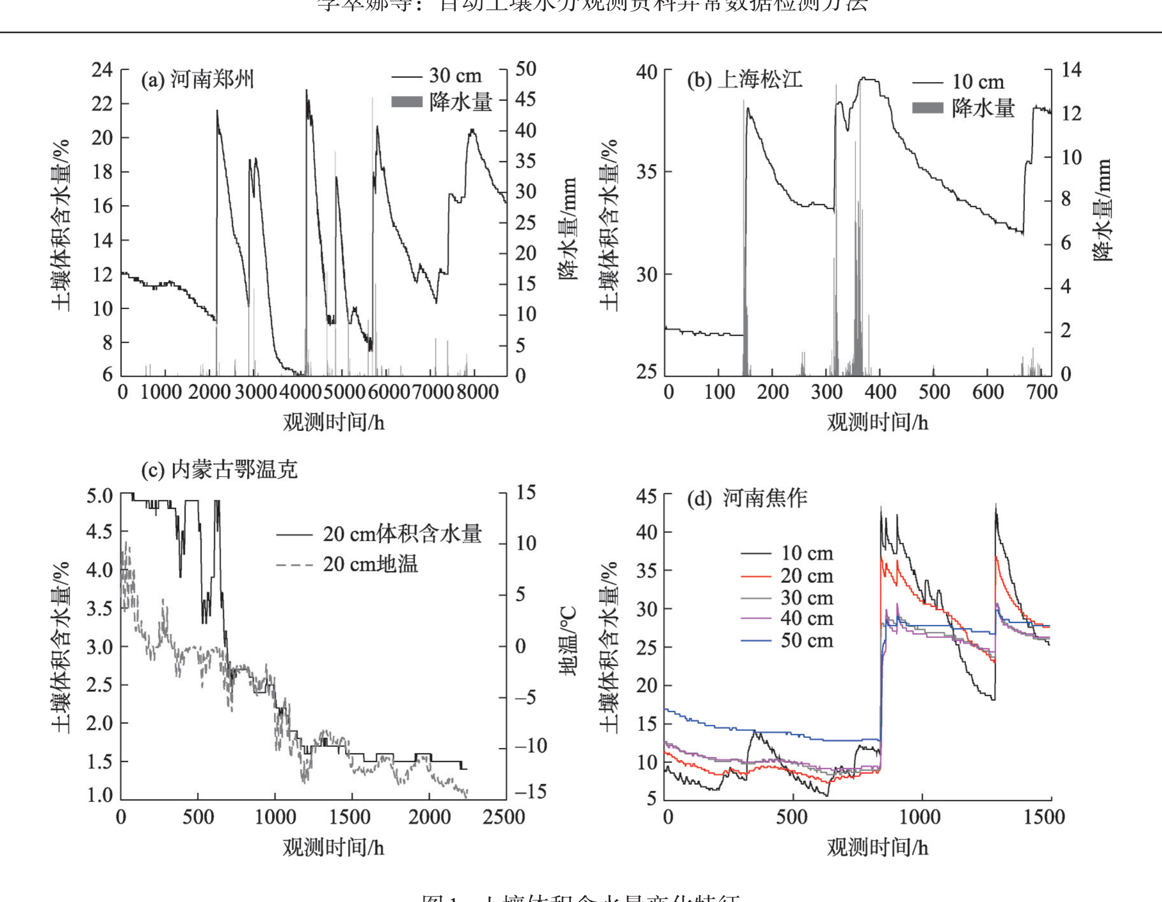


图1　土壤体积含水量变化特征

Fig. 1　The variation characteristics of soil volumetric water content

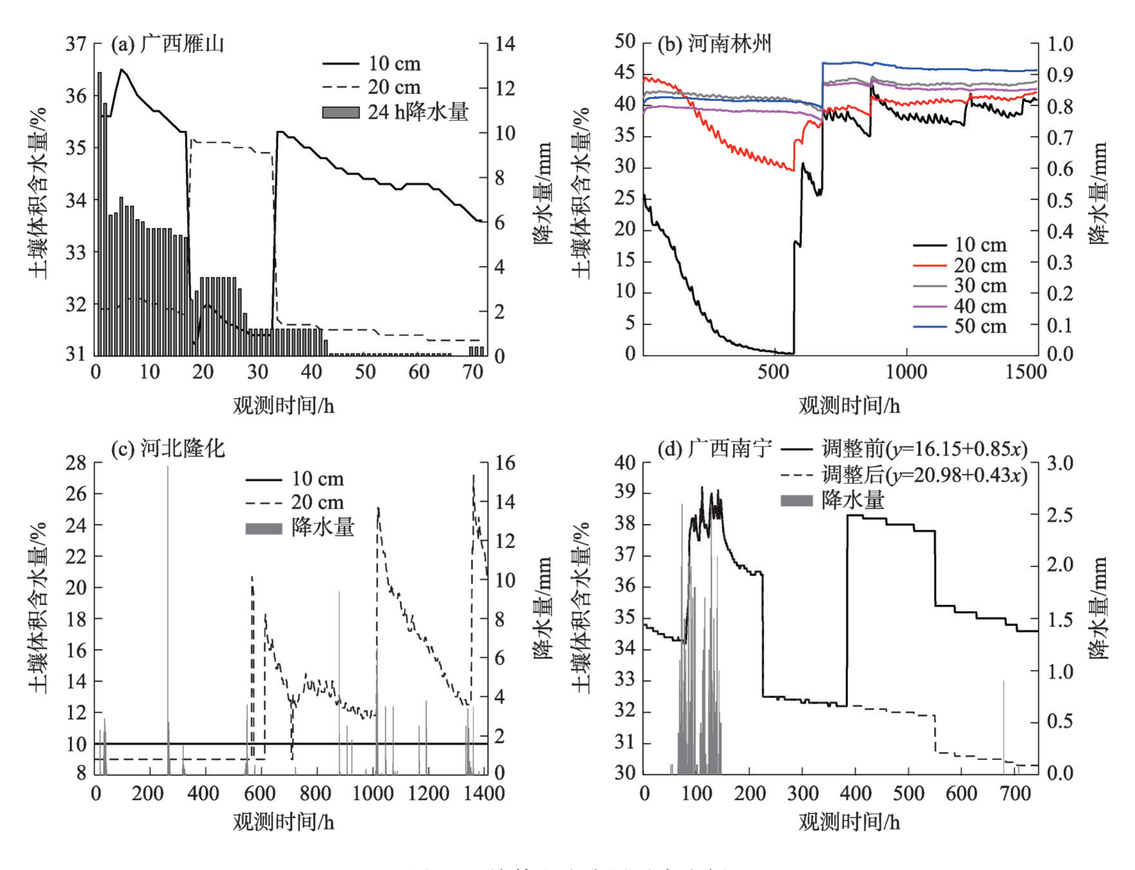


图2　土壤体积含水量异常个例

Fig. 2　Examples of anomalous soil volumetric water content

式中：$f_{tD\min}$ 和 $f_{tD\max}$ 分别为 t 时刻 D 深度的土壤传感器在水和空气中所测的频率值；f_{tD} 为 t 时刻 D 深度的土壤传感器的测量值。

2.3.2 体积含水量和相对湿度界限值检测（VD & RHD） 依据土壤水分传感器测量范围，结合土壤质地的保水性特征，以保水性最好的黏土在过饱和时所能达到的体积含水量为界限值设定阈值（公式2）；但因灌溉、强降水等因素造成土壤的过饱和以及田间持水量参数测定偏小造成土壤相对湿度短时间超过100%现象，以三倍标准差设定阈值[22-24]（公式3），超过该范围为粗大误差，数据错误。

$$\text{VD:}\ 0 < X_{tD} < 60 \quad (2)$$

$$\text{RHD:}\ 0 < Y_{tD} < 180 \quad (3)$$

式中：VD为体积含水量阈值检测；RHD为相对湿度阈值检测；X_{tD} 为 t 时刻 D 深度的土壤体积含水量值；Y_{tD} 为 t 时刻 D 深度的土壤相对湿度值。

内蒙古赛罕区2019年9月第二层土壤体积含水量异常（图3a中红色标识），经与台站核查为仪器标定方法不适用，重新标定后体积含水量观测值恢复正常（图3a中虚线所示）；青海边滩（X1811）2019年8月第一层土壤相对湿度异常（图3b中红色标识），相对湿度在150%~350%范围波动，二者均随降水有明显的正响应关系，初步诊断为土壤水文物理常数测定不准确，经参数调整，相对湿度从300%的异常水平恢复到50%的正常水平。

2.3.3 无降水突变（SBD） 降水是影响土壤水分变化的重要变量，无降水突变是基于二者间的一致性关系而检测出异常值的方法。在某 t 时刻，如果24 h累计降水量小于等于某一临界值，且满足公式（3）和（4）或（5）条件，即为异常值。

$$X_{tD} > X_{(t-1)D} \quad (4)$$

$$X_{tD} - X_{(t-24)D} > 2s_{x(t-24t)D} \quad (5)$$

$$|X_{tD} - X_{(t-1)D}|^3 > 3 \quad (6)$$

式中：$X_{(t-1)D}$ 为 $t-1$ 时刻 D 深度的土壤体积含水量值；$X_{(t-24)D}$ 为 $t-24$ 时刻 D 深度的土壤体积含水量值；$s_{x(t-24t)D}$ 为过去24 h内深度为 D 的土壤体积含水量标准偏差。24 h累计降水量临界值与观测深度和

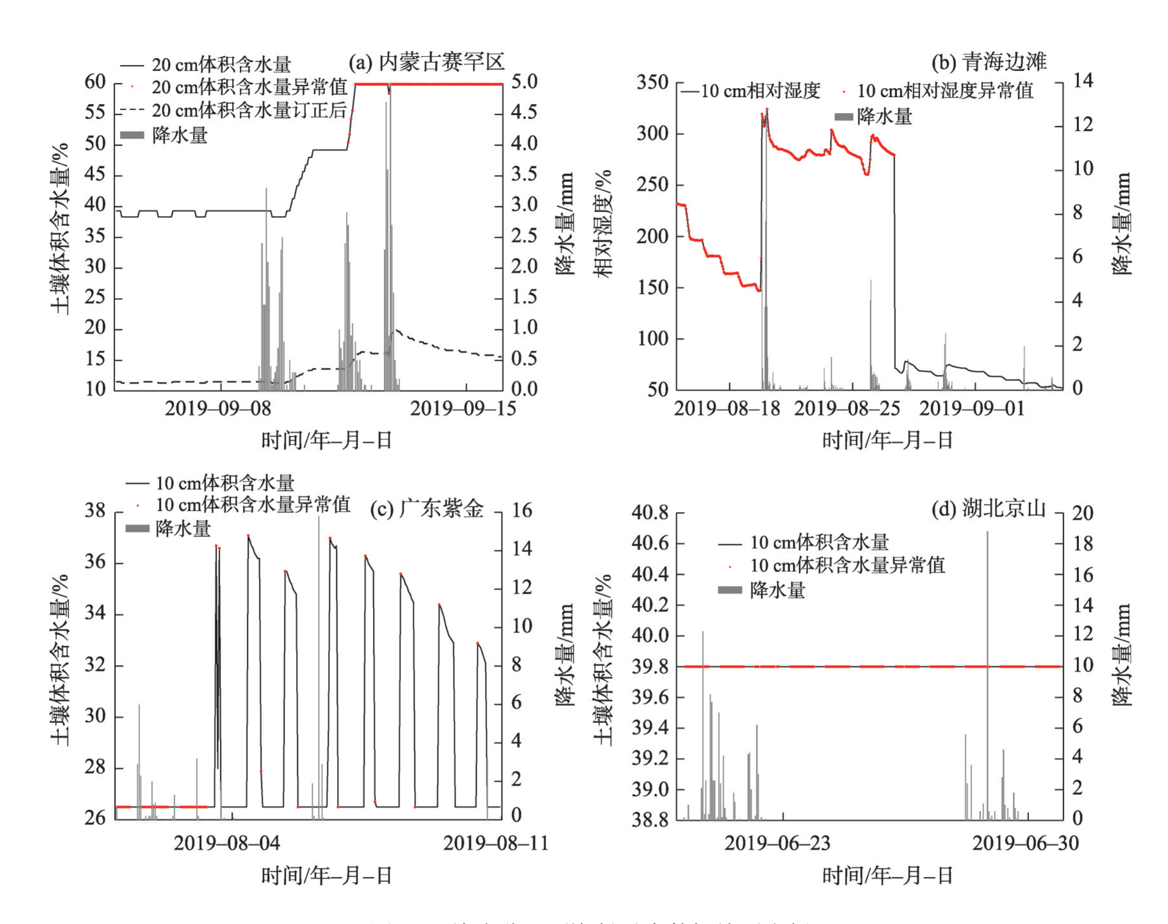


图3 土壤水分观测资料异常数据检测个例

Fig. 3 Examples of abnormal data detection in soil moisture observation data

传感器精度有关，可用公式(7)表示。

$$P_{min} > \frac{DAp}{100} \tag{7}$$

式中：D指传感器的观测深度(cm)；A是传感器精度；p是土壤孔隙度。A、p通常取值为0.05 $m^3 \cdot m^{-3}$和0.5。基于降水的土壤体积含水量异常值检测更适用于表层(0~10 cm)。

2.3.4 异常峰值检测(SD) 改进传统的峰值检测方法将解决降水引起的土壤含水量升高标记为错误的问题。基于土壤体积含水量值时序变化以及二次导数的峰值检测算法识别异常值。若t时刻深度为D的观测值较前一时刻增加或减少了至少15%，则该时刻的观测值是一个潜在的异常峰值。

$$\frac{X_{tD}}{X_{(t-1)D}} > 1.15 \text{ 或 } \frac{X_{tD}}{X_{(t-1)D}} < 0.85 \tag{8}$$

增加二阶导数对上述潜在异常峰值进行进一步检测，t-1和t+1时刻的土壤体积含水量二阶导数($X''_{(t-1)D}$、$X''_{(t+1)D}$)的比率介于0.8~1.2之间，反之异常，即

$$0.8 < \left| \frac{X''_{(t-1)D}}{X''_{(t+1)D}} \right| < 1.2 \tag{9}$$

其中，二阶导数采用Savitzky-Golay 3 h平滑求导和二项多项式拟合求得，即

$$X''_{tD} = \frac{1}{7}[2\left(X_{(t-2)D} + X_{(t+2)D}\right) - \left(X_{(t-1)D} + X_{(t+1)D}\right) - 2X_{tD}] \tag{10}$$

由于二阶导数不适用于随机噪声，在满足上述条件基础上增加检测t时刻前后12 h内土壤体积含水量的平均值(m)与方差(S^2)的关系，即

$$\left| \frac{S^2(X_{(t-12)D},\ X_{(t+12)D})}{m(X_{(t-12)D},\ X_{(t+12)D})} \right| < 1 \tag{11}$$

如果体积含水量观测值同时满足公式(8)–(11)，则该值为异常峰值。

广东紫金站2019年8月第一层土壤体积含水量异常(图3c中红色标识)，体积含水量随降水无明显的正响应关系，观测值在26.5%~36.5%间呈锯齿状变化，与土壤水分变化特征不符。

2.3.5 恒值检测(CD) 恒值检测也称最小变率检测，由于观测仪器发生故障、霜冻等原因造成观测值长时间不变或微变，引起观测记录不真实；为区别降水造成的土壤水分正常过饱和，需持续达到一定时间，土壤体积含水量的变化小于传感器精度的1%，即0.0005 $m^3 \cdot m^{-3}$，见公式(12)–(13)。

10~20 cm可用48 h内最高值和最低值检测恒值，即

$$X_{max(48\ h)} - X_{min(48\ h)} < 0.0005 \tag{12}$$

30~50 cm可用15 d内最高值和最低值检测恒值，即

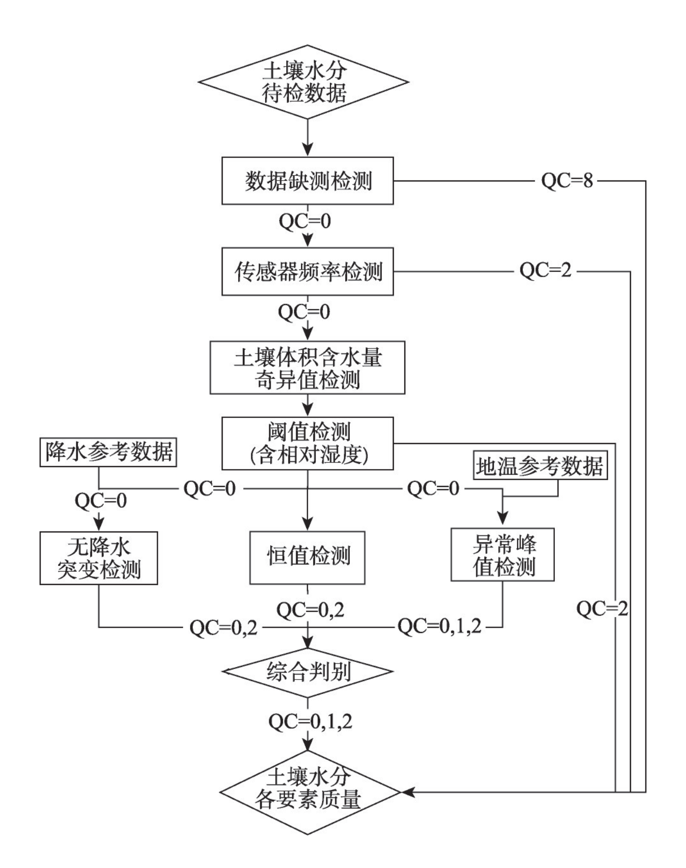
$$X_{max(15\ d)} - X_{min(15\ d)} < 0.0005 \tag{13}$$

湖北京山站(57387)2019年6月第一层土壤体积含水量异常(图3d)，长时间维持在39.8%，与土壤水分变化特征不符。

2.4 检测流程

本文借鉴ISMN提出的谱分析方法思路[10]，制定土壤水分数据质量控制和综合分析流程，同时参照地面气象观测资料质量控制标识[25]设计土壤水分观测数据质量标识，异常数据检测流程见图4。

待检数据缺测或缺报，缺测层各要素质控码标记为8，不再进行其他检查；待检数据频率未通过检查，相应层传感器故障、各要素判断错误，质控码标



注：QC为质量控制码。

图4 土壤水分自动观测异常数据检测流程

Fig. 4 Abnormal data detection process of automatic soil moisture observation

记为2，不再进行其他检查；体积含水量和相对湿度未通过阈值检查，质控码标记为2，相应层观测数据质控码标记为2，不再进行其他检查。

无降水突变检测、异常峰值检测、恒值检测三项检查中，任意一项未通过检查，判定为错误或可疑，则体积含水量质量标记为2或1，三项检查均正确，质量标记为0，其他要素采用与体积含水量相同的标记。

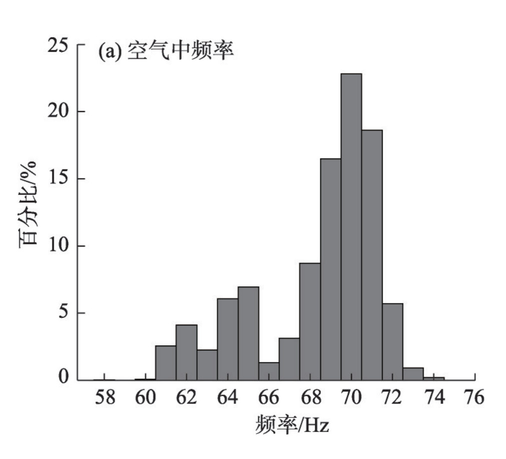
3 土壤水分观测资料检验与分析

3.1 基于频率特征的传感器故障检测

依据全国725个土壤水分自动站5295支传感器在空气和水中频率的分布特征（图5），传感器在水中频率值主要集中在41~44之间，占71.90%；传感器在空气中频率值主要集中在67~72之间，占72.31%。由于不同地区土壤中有机质含量、酸碱度等不同，频率界限略有差异。

以2015年新疆维吾尔自治区34个自动土壤水分站数据为检验对象，表1给出了频率值检验（FD）、体积含水量值检验（VD）结果对比。FD异常值检出量为64817个，约占7.4%，VD异常值检出量为38537个，约占4.4%，经核定福海等站检出的异常值全部是由传感器故障引起的，FD法能更好的定位到数据质量问题源。

分析异常数据中FD异常值和VD异常值的数量（图6）。新疆北部和中部等7站FD异常值数量明显较其他站点多，新疆西部站点异常值则普遍较少，因此，FD检出异常值较少的站点有效数据质量一般较高，而FD检出的异常值主要由传感器故障引起。各站点VD检出的异常值一般少于5000个，但炮台站的VD异常值检出量达到129700个，阜康与岳普湖的VD异常值检出量也较高。



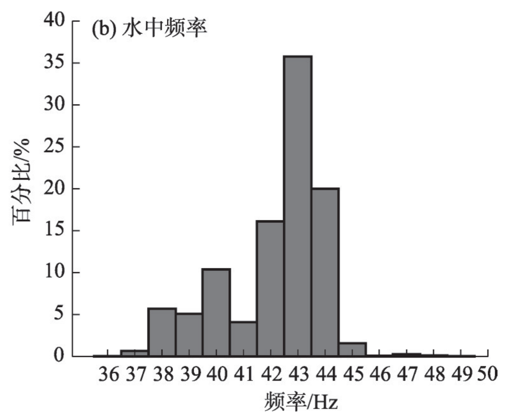


图5 土壤水分传感器在空气和水中频率分布特征

Fig. 5 Frequency distribution characteristics of soil moisture sensor in air and water

表1 2种方法异常值检验结果对比

Tab. 1 Comparison of abnormal data detection results by three methods

指标	VD	FD
异常值数量/个	38537	64817
百分比/%	4.4	7.4

注：VD为体积含水量值检验；FD为频率值检验。

3.2 基于体积含水量的异常值检验分析

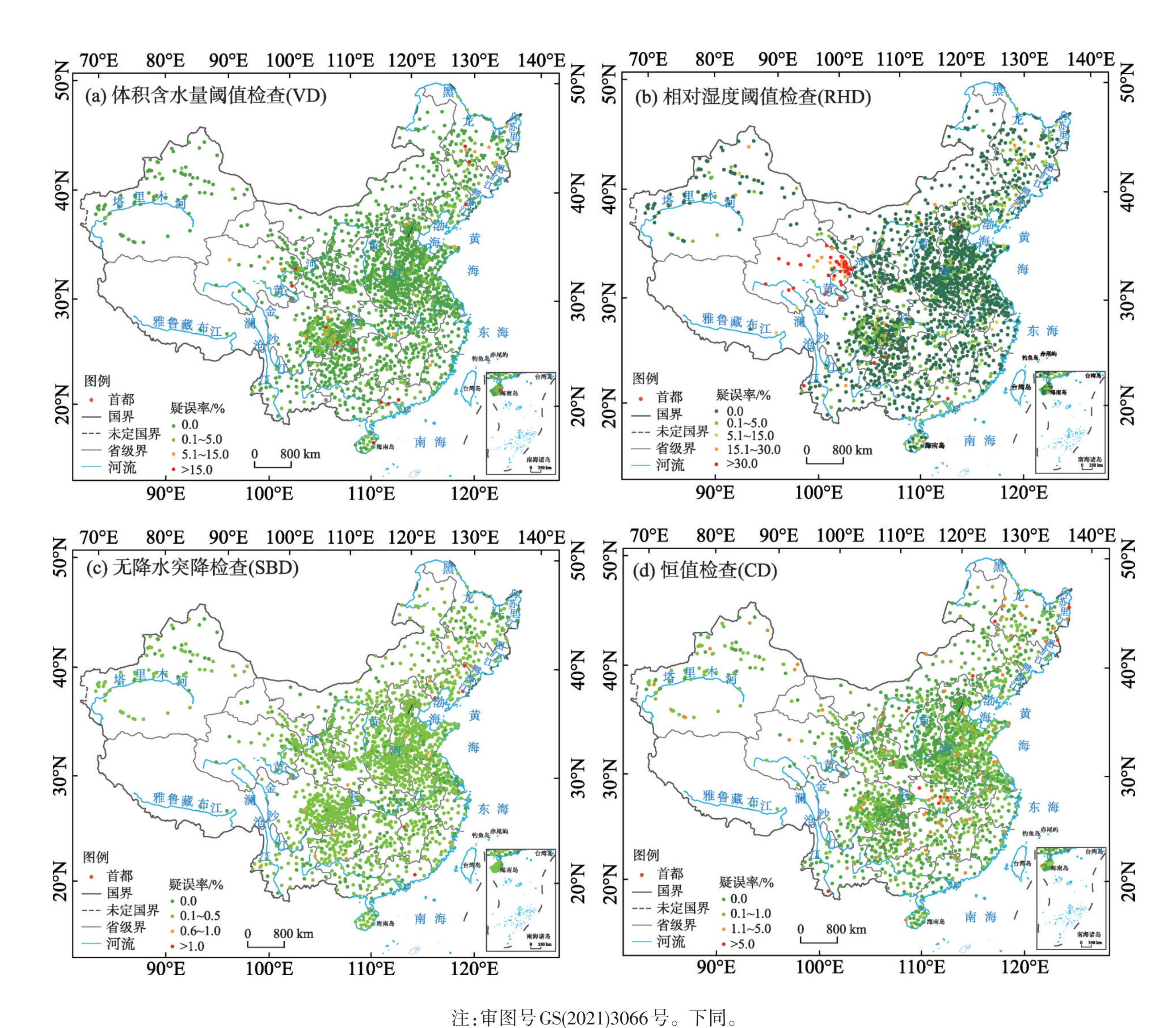
应用第2节异常数据检测方案和流程，对2019年6—9月的2088个自动土壤水分站小时资料进行了检测，数据正确率为93.8%，疑误率为1.7%，缺测率为4.5%。其中，VD、RHD、SBD和CD异常数据检出率分别为0.3%、1.1%、0.1%和1.7%。

图7给出4种方法检测的土壤水分观测数据质量空间分布情况。VD和RHD异常值检出量相对较高，且以少量站点长时间质量问题为主。其中，RHD异常值检出量最高达48×10^{5}条，占全网观测站点10.5%，以青海省问题最为突出。该类质量问题主要由人工土壤水文物理常数测定不准引起。VD异常值检出量次之，为11×10^{5}条，占全网观测站点13.9%，主要分布在黑龙江、青海和广东等地。该类质量问题主要由设备故障引起。SBD和CD异常值检出量相对较低，分别为3×10^{5}条和1×10^{5}条，占全

注:该图基于国家测绘地理信息局标准地图服务网站下载的审图号为GS(2017)3320号的标准地图制作,底图无修改。

图6　2015年新疆维吾尔自治区自动站土壤水分数据频率检测和体积含水量检测异常值

Fig. 6　Abnormal values were detected by frequency and volumetric water content of Xinjiang Uygur Autonomous Region in 2015



注:审图号GS(2021)3066号。下同。

图7　异常站点的空间分布

Fig. 7　Spatial distribution of abnormal sites

网观测站点86.7%和48.1%。上述两类质量问题主要由传感器标定或设备性能下降等引起。

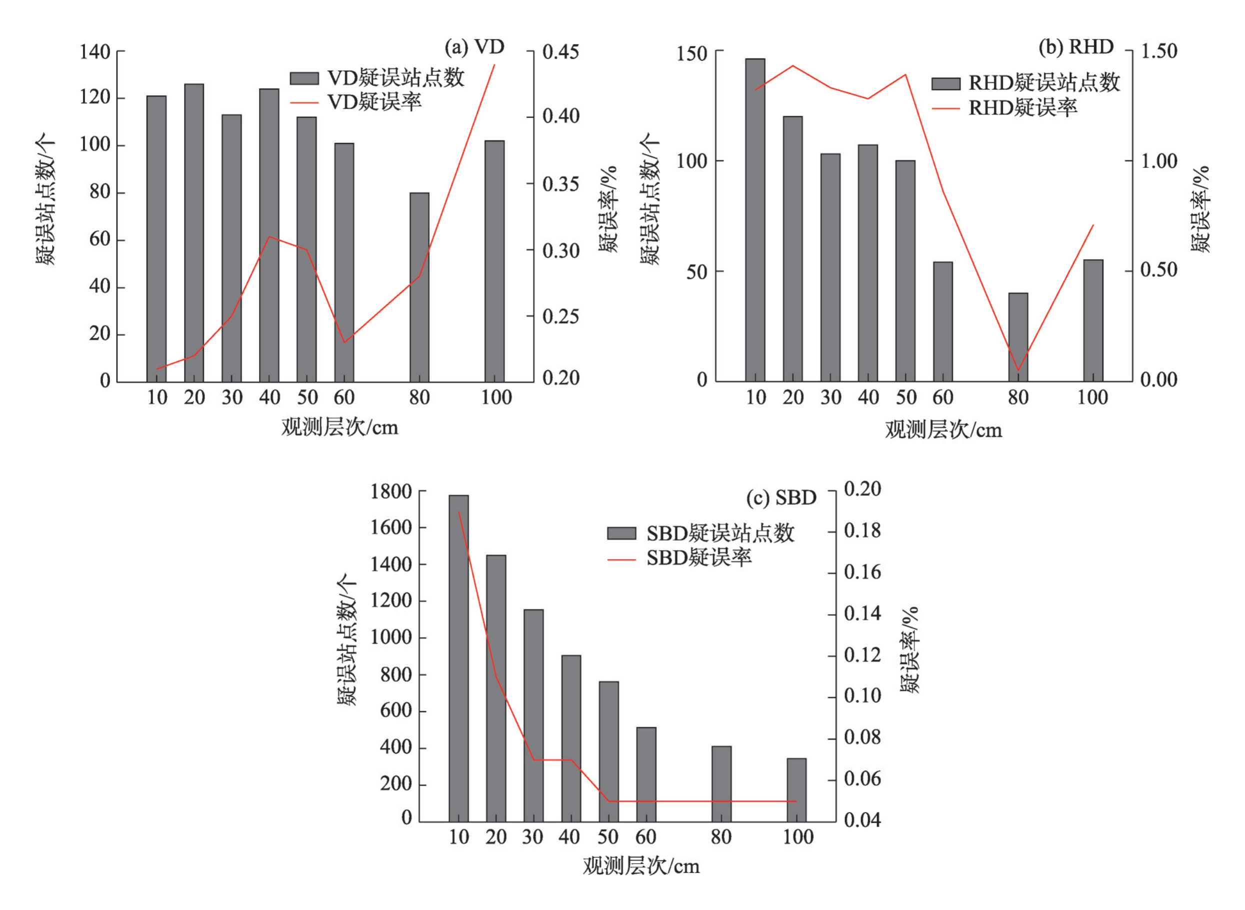
图8a~c给出VD、RHD和SBD 3种方法检测的土壤水分异常值层次分布情况。土壤水分异常值主要集中在浅层和中层(0~50 cm)。VD异常值检出站数各层基本持平,在80~126范围波动,但疑误率随观测深度呈先增后降再升的变化趋势,100 cm疑误率最高为0.44%,主要出现在四川、广东等省,初步分析可能与下垫面或地下水位有关(图8a)。各层RHD异常值检出站点数随观测深度呈下降趋势且波动范围较大,在40~146变化,0~50 cm观测层疑误率保持平稳,基本在1.28%~1.43%范围波动(图8b)。SBD异常值检出站数与疑误率均随观测深度呈明显下降趋势,10 cm疑误率最高0.19%,这可能是土壤表层土壤水分易受外部环境(降水、蒸发等)影响造成(图8c)。图9给出了基于VD、RHD和SBD 3种方法整个观测层均检出异常值的站点分布情况。从空间分布看,主要集中在新疆、青海、四川、河南和贵州等地,为提醒装备保障人员及时进行设备标定和维护提供技术支撑。

4 结 论

为剔除自动土壤水分观测资料中的异常数据,本文提出了一套适用于自动土壤水分观测中的异常数据检测方法。以历史自动土壤水分站观测资料,研究自动土壤水分观测资料固有变化特征,结合仪器观测原理、数据特征和异常数据误差来源,分析异常数据并制定质控方法,采用2019年全国土壤水分观测资料对异常数据检测方法应用效果进行检验。结果表明:

(1) 基于频率特征仪器故障检测方法,通过引入传感器直接测量值在空气和水中频率值变化特征,可有效检测出因仪器故障引起的错误数据,准确率达100%。

(2) 结合土壤水分观测中的突出质量问题,基于土壤水分观测资料固有特征及异常数据误差来源,设计的VD、RHD、SBD和CD四类检测方法可有



注:VD为体积含水量阈值检查;RHD为相对湿度阈值检查;SBD为无降水突降检查。下同。

图8 异常数据各层次分布

Fig. 8 Level distribution of abnormal data

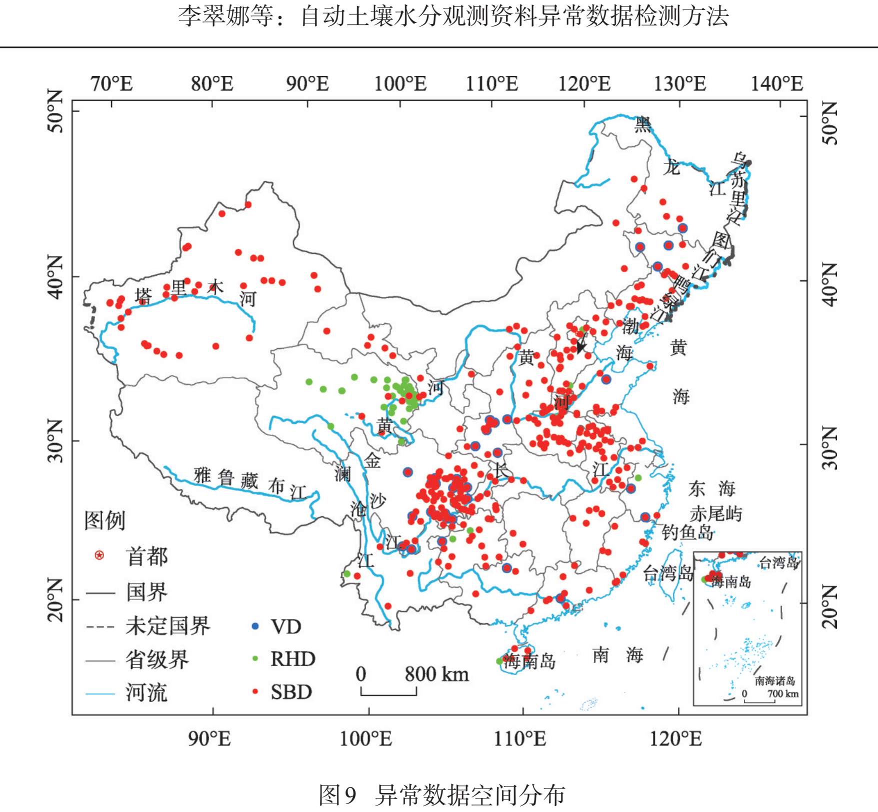


图9 异常数据空间分布

Fig. 9 Spatial distribution of abnormal data

效检测出土壤水分观测中的异常数据，可帮助用户快速发现并解决观测端引起的质量问题。

（3）2019年6—9月的全国自动土壤水分站小时资料数据正确率为93.8%，疑误率为1.7%，缺测率为4.5%；异常数据主要分布在青海、四川、山东、河北和广东等地，土壤表层（0~20 cm）疑误率较高，少数站点出现整个观测层异常情况；全国大部分土壤水分自动站存在数据质量异常现象，以少数站点存在长期性质量问题为主，主要由传感器标定或土壤水文物理常数测定不准等原因造成。

（4）该方法已初步应用于国家级综合气象观测数据质量控制业务系统（天衡），为进一步提高该方法的适用性和准确性，需针对土壤水分实际观测中各种异常数据以及持续一段时间的异常峰值和突降数据进行自动检测方法研究，以期进一步提高土壤水分数据质量。

参考文献（References）

[1] 韩姣姣, 段旭, 赵洋毅. 金沙江干热河谷不同植被坡面土壤水分时空分布特征[J]. 干旱区地理, 2019, 42(1): 121−129. [Han Jiaojiao, Duan Xu, Zhao Yangyi. Spatial and temporal variability of soil moisture on slope land of different vegetation of dry-hot valley in Jinsha River[J]. Arid Land Geography, 2019, 42(1): 121−129.]

[2] 蔡亮红, 丁建丽. 基于改进植被指数土壤水分遥感反演[J]. 干旱区地理, 2017, 40(6): 1248−1255. [Cai Lianghong, Ding Jianli. Remote sensing inversion of soil moisture based on modified vegetation index[J]. Arid Land Geography, 2017, 40(6): 1248−1255.]

[3] 丁新原, 周智彬, 雷加强, 等. 塔里木沙漠公路防护林土壤水分特征曲线模型分析与比较[J]. 干旱区地理, 2015, 38(5): 985−993. [Ding Xinyuan, Zhou Zhibin, Lei Jiaqiang, et al. Analysis and comparison of models for soil water characteristic curves of Tarim Desert Highway Shelterbelt[J]. Arid Land Geography, 2015, 38(5): 985−993.]

[4] 陈丙寅, 杨辽, 陈曦, 等. 基于改进型TVDI在干旱区旱情监测中的应用研究[J]. 干旱区地理, 2019, 42(4): 902−913. [Chen Bingyin, Yang Liao, Chen Xi, et al. Application of modified TVDI in drought monitoring in arid areas[J]. Arid Land Geography, 2019, 42(4): 902−913.]

[5] 王新友, 马全林, 靳虎甲, 等. 石羊河下游人工梭梭林土壤呼吸变化特征及其与水热因子的关系[J]. 干旱区地理, 2019, 42(3): 570−580. [Wang Xinyou, Ma Quanlin, Jin Hujia, et al. Soil respiration variation characteristics and its relationship with hydrothermic factor of artificial *Haloxylon ammodendron* forest in lower reaches of Shiyang River[J]. Arid Land Geography, 2019, 42(3): 570−580.]

[6] 王钧, 李广, 聂志刚, 等. 陇中黄土丘陵沟壑区人工草地土壤水蚀预测模型[J]. 干旱区地理, 2020, 43(2): 398−405. [Wang Jun, Li Guang, Nie Zhigang, et al. Testing a soil water erosion predictive model in an artificial grassland in the hill gully Loess Plateau [J]. Arid Land Geography, 2020, 43(2): 398−405.]

[7] 张志富. 自动站土壤水分资料质量控制方案的研制[J]. 干旱区地理, 2013, 36(1): 101−108. [Zhang Zhifu. Quality control schemes

for hourly soil moisture from automatic weather stations in China [J]. Arid Land Geography, 2013, 36(1): 101–108.]

[8] 王晓东, 杨太明, 吴必文, 等. 安徽省自动观测土壤水分质量控制方法[J]. 气象科技, 2015, 43(3): 399–404. [Wang Xiaodong, Yang Taiming, Wu Biwen, et al. Error analysis and quality control method of automatic soil moisture measurements in Anhui[J]. Meteorological Science and Technology, 2015, 43(3): 399–404.]

[9] 王良宇, 何延波, 张艳红, 等. 一种自动土壤水分数据质量控制预警方法[J]. 气象科技, 2016, 44(4): 528–534. [Wang Liangyu, He Yanbo, Zhang Yanhong, et al. Data quality early warning of automatic soil water observations based on soil hydrological constants [J]. Meteorological Science and Technology, 2016, 44(4): 528–534.]

[10] 吴东丽, 曹婷婷, 薛红喜. 自动土壤水分观测数据质量控制方法及其应用[J]. 土壤科学, 2016, 4(1): 1–10. [Wu Dongli, Cao Tingting, Xue Hongxi. The study of quality control for observing data of automatic soil moisture[J]. Hans Journal of Soil Science, 2016, 4(1): 1–10.]

[11] 王佳强, 赵煜飞, 任芝花, 等. 中国自动土壤水分观测资料质量控制方法设计与效果检验[J]. 气象, 2018, 44(2): 244–257. [Wang Jiaqiang, Zhao Yufei, Ren Zhihua, et al. Design and verification of quality control methods for automatic soil moisture observation data in China[J]. Meteorological Monthly, 2018, 44(2): 244–257.]

[12] Dorigo W A, Xaver A, Vreugdenhil M, et al. Global automated quality control of in situ soil moisture data from the International Soil Moisture Network[J]. Vadose Zone Journal, 2013, 12(3): 918–924.

[13] 王良宇, 何延波. 自动土壤水分观测数据异常值值域研究[J]. 气象, 2015, 41(8): 1017–1022. [Wang Liangyu, He Yanbo. Research on outlier threshold of automatic soil moisture observation data[J]. Meteorological Monthly, 2015, 41(8): 1017–1022.]

[14] 中国气象局. 自动土壤水分观测规范(试行)[R]. 北京: 中国气象局, 2010. [China Meteorological Administration. Specification for automatic soil moisture observation(trial) [R]. Beijing: China Meteorological Administration, 2010.]

[15] 吴芹芹, 莫淑红, 程圣东, 等. 黄土区冻融期不同土地利用土壤水分与温度的关系[J]. 干旱区研究, 2020, 37(3): 627–635. [Wu Qinqin, Mo Shuhong, Cheng Shengdong, et al. Study on the correlation between soil moisture and temperature of different land uses in the loess area during a freezing-thawing period[J]. Arid Zone Research, 2020, 37(3): 627–635.]

[16] 裴步祥, 毛飞. 我国土壤水分的变化规律和土壤湿度测站合理布局的研究[J]. 北京农业大学学报, 1990, 16(增刊1): 1–12. [Pei Buxiang, Mao Fei. Study on soil moisture variation patterns and the rational distribution of soil moisture observation stations of China[J]. Acta Agriculture Universities Pekinensis, 1990, 16 (Suppl. 1): 1–12.]

[17] 吕明侠, 王一博, 刘国华. 黑河上游高山区浅层土壤水分变化及其影响因素[J]. 干旱区研究, 2020, 37(4): 899–907. [Lyu Mingxia, Wang Yibo, Liu Guohua. Changes in shallow soil moisture and its influencing factors in the alpine region of the upper Heihe River [J]. Arid Zone Research, 2020, 37(4): 899–907.]

[18] Zhao Chunlei, Shao Ming'an, Jia Xiaoxu, et al. Spatial distribution of water-active soil layer along the south-north transect in the Loess Plateau of China[J]. Journal of Arid Land, 2019, 11(2): 228–240.

[19] Gong Yidan, Xing Xuguang, Wang Weihua. Factors determining soil water heterogeneity on the Chinese Loess Plateau as based on an empirical mode decomposition method[J]. Journal of Arid Land, 2020, 12(3): 462–472.

[20] Basara J B, Crawford T M. Improved installation procedures for deep-layer soil moisture measurements[J]. Journal of Atmospheric Ocean Technology, 2000, 17(6): 879–884.

[21] You J S, Hubbard K G, Mahmood R, et al. Quality control of soil water data in applied climate information system: Case study in Nebraska[J]. Journal Hydrologic Engineering, 2010(15): 200–209.

[22] Hubbard K G, Goddard S, Sorensen W D, et al. Performance of quality assurance procedures for an applied climate information system[J]. Journal of Atmospheric and Oceanic Technology, 2005 (22): 105–112.

[23] Gonzálezrouco J F, Jiménez J L, Quesada V, et al. Quality control and homogentity of precipitation data in the douthwest of Europe [J]. Journal of Climate, 2001, 14(14): 964–978.

[24] Journée M, Bertrand C. Quality control of solar radiation data within the RMIB solar measurements network[J]. Solar Energy, 2011, 85: 72–86.

[25] 中国气象局. QXT118–2010. 地面气象资料质量控制标准[S]. 北京: 气象出版社, 2010. [China Meteorological Administration. QXT118–2010. Quality control of surface meteorological observational data[S]. Beijing: China Meteorological Press, 2010.]

Anomaly data detection method for in situ automatic soil moisture observation

LI Cuina[1], LIU Tianqi[2], WU Dongli[1]

(1. Meteorological Observation Centre, CMA, Beijing 100081, Beijing, China; 2. Inner Mongolia Meteorological Information Center, CMA, Hohhot 010051, Inner Mongolia, China)

Abstract: Soil moisture plays a crucial role in the study of agricultural drought monitoring, yield prediction, and soil erosion, which are of great significance for agriculture, drought, and climate studies. Automatic soil moisture observation instruments have become an important component of the automatic soil moisture observation stations run by the meteorological department in China given their high precision, high spatial and temporal resolution, and nondestructive capabilities. The accuracy of data obtained through the process of observing soil moisture is seriously affected by the calibration methods used, the stability of the equipment, and the texture of the soil. Thus, it is especially important to establish a quality control method for automatic soil moisture observation data that are free from these error sources that affect the observation quality. In attempting to solve the outstanding quality problems in automatic soil moisture observation data, this paper studies the inherent variation characteristics of soil moisture on the basis of the data obtained by the automatic soil moisture observation stations in China between 2013 and 2015. Combined with the instrument observation principle and the data characteristics and error sources of abnormal data, this study classifies and statistically analyzes the form of the abnormal data, and given the threshold, it puts forward a preliminary practical set of quality control methods for the hourly soil moisture observation data, which includes frequency detection (FD), threshold detection for volume moisture content and relative humidity, break drop detection, sudden change detection, and constant detection. The application effect of this quality control method has been verified using data obtained in China in 2019. The results show that (1) the four main categories of abnormal data are gross errors, mutations, abnormally high values, and stiffness; these are mainly caused by instrument failure, abnormal soil hydrological constants, and unreasonable calibration of sensors. (2) Based on the FD method, which analyzes the change characteristics through the frequency values measured by the sensors in the air and water, this quality control method can effectively detect the abnormal data caused by instrument failure, with the accuracy rate reaching 100%. (3) The evaluation results from the hourly data of the national automatic soil moisture stations from June to September 2019 show that the accuracy rate of data is 93.8%, the abnormal rate is 1.7%, and the missing rate is 4.5%. The abnormal data are mainly distributed in the Qinghai, Sichuan, Shandong, Hebei, and Guangdong Provinces. The soil surface layer (0−20 cm) has a higher proportion of anomalous data than that of other layers, and only a few stations have an anomaly within the whole observation layer. Although the abnormal phenomena can be observed at most sites across the country, the primary problem is that a small number of sites have long-term abnormalities, which have been caused either by the calibration of the sensor or the inaccurate determination of the soil's hydrological and physical constants. (4) This quality control method can effectively detect abnormal data in China. Presently, it has been applied to the Integrated Meteorology Observation Data Quality Control System.

Key words: automatic soil moisture observation; quality control; abnormal data; instrument observation principle